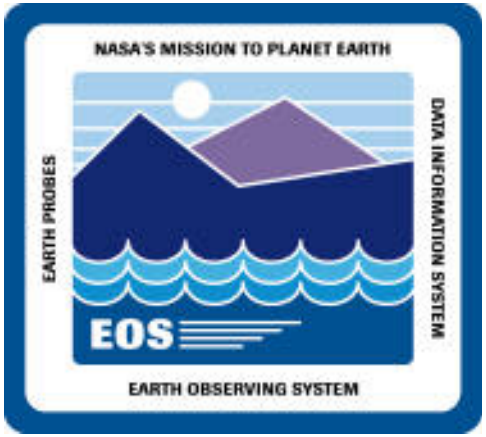


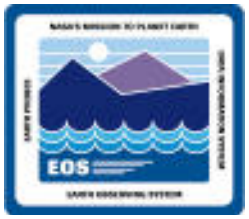
Production Subsetting

William Knauss

wknauss@eos.hitc.com

16 April 1996

PDPS Roadmap

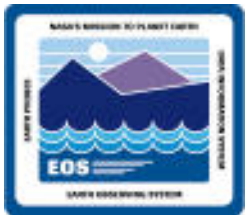


Special Topic: Production Rules
Capture PGE Profile at **SSI&T**
Describe Production Goals through **Production Requests**
Accept **On-demand** Production Requests
Accept Resource Reservations and Create **Resource Plans**
Planning Production Controls - Create and Activate **Production Plans**
Coordinate Production from Data Arrival with **Subscription** Notifications
Handle L0 **Data Preparation**

Special Topic: Production Subsetting

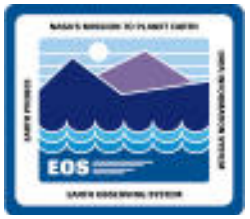
Realtime **Production** Controls and PGE Execution Monitoring
Special Topic: PGE Exit Handling
Quality Assurance Check Output Products
Special Topic: PDPS Database
Special Topic: Ancillary Data Pre-Processing

Overview



- Purpose
- Types of Production Subsetting
- Spectral Subsetting
- Swath-Based Subsetting
- Geographic Masking
- Compression
- Summary

Purpose of Production Subsetting



Background:

- **Inter-DAAC data transfers required for routine processing and reprocessing place a heavy load on the network. Primary bottlenecks are the GSFC to EDC and GSFC to LaRC transfers of MODIS data.**

Approach:

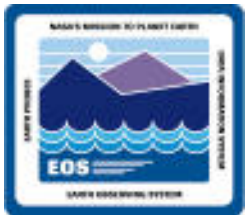
- **Subset the data prior to transfer so that only the data required at a remote DAAC is sent over the network.**

A detailed analysis of this topic can be found in the Technical Paper *Reducing Inter-DAAC Data Transfers Through Subsetting*; 160-TP-005-001.

Types of Subsetting

- **Spectral Band Subsetting**
- **Swath-based Subsetting**
- **Geographical Masking and Compression**
- **Compression for Fill Values**

Level 3 Requirements



SDPS-0200#B - In support of reducing production data dependency flow bandwidth consumption during inter-DAAC network transmission, the ECS shall support subsetting through the use of geographical masking (land/sea mask, snow/ ice mask) for standard production as well as reprocessing.

SDPS-0210#B - In support of reducing production data dependency flow bandwidth consumption during inter-DAAC network transmission, the ECS shall support the application of lossless compression and decompression techniques on data set files for removal of data set fill pixels, for standard production as well as reprocessing.

SDPS-0220#B - In support of reducing production data dependency flow bandwidth consumption during inter-DAAC network transmission, the ECS shall support subsetting through swath width reduction by selection of a range of pixels from each swath row, for standard production as well as reprocessing.

SDPS-0230#B - In support of reducing production data dependency flow bandwidth consumption during inter-DAAC network transmission, the ECS shall support subsetting by spectral band(s) selection, for standard production as well as reprocessing.

Spectral Subsetting

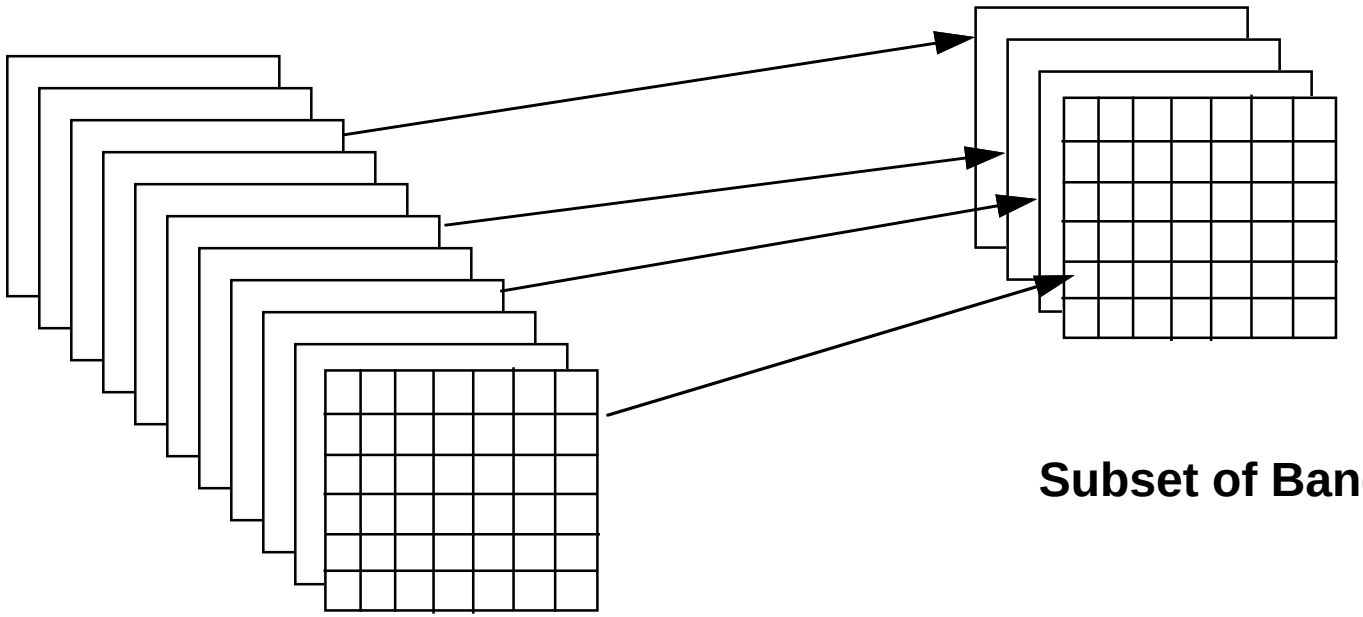
Background:

- **Only certain spectral bands of a data set may be required at a remote DAAC. For example, only 15 of 36 MODIS Level 1B spectral bands are required for CREEPS and MISR processing at LaRC.**

Approach:

- **Spectral subsetting is directly supported by the Data Server and HDF-EOS. The DAAC requiring data will simply request the bands needed for processing.**

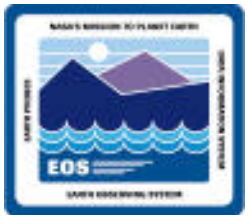
Spectral Subsetting



Subset of Bands

Original Bands

Swath-Based Subsetting



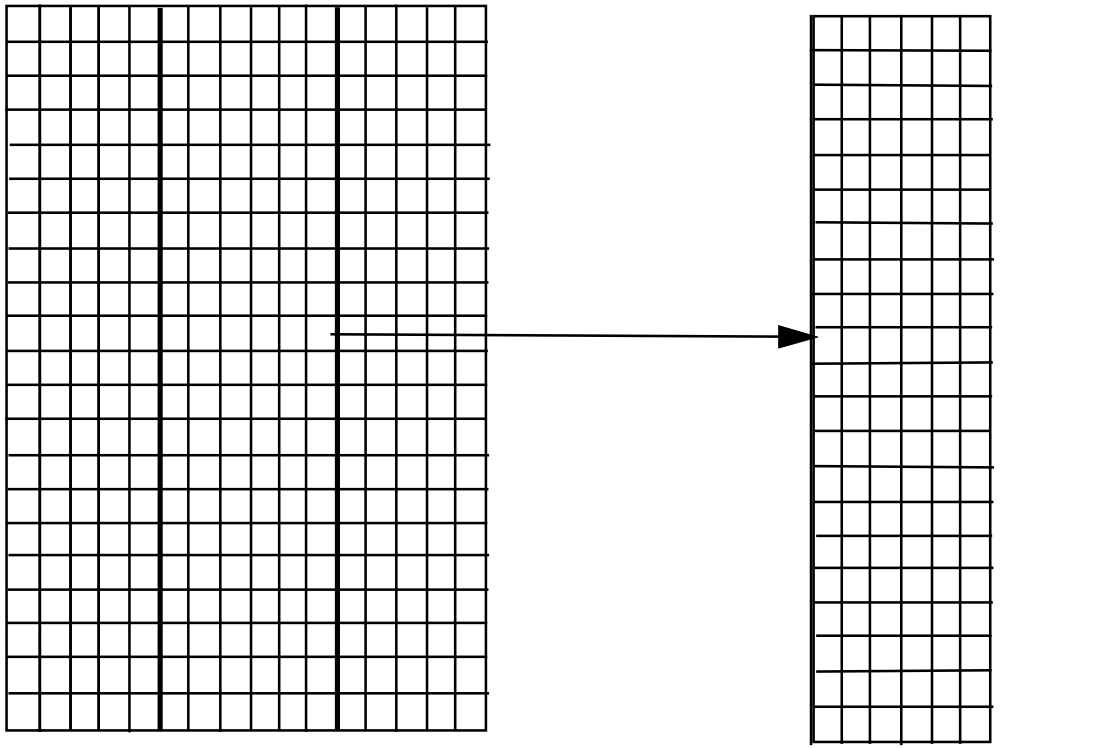
Background:

- **MISR and MOPITT processing at LaRC uses MODIS swath data, but only requires a approximately 400 of the 1354 pixels of the MODIS swath.**

Approach:

- **Swath-based subsetting will be performed by running a PGE at the originating DAAC. The subsetted data will then be transferred to the receiving DAAC for subsequent processing.**

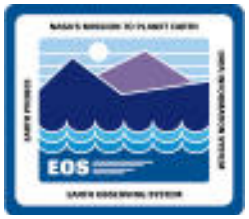
Swath-Based Subsetting



Full Swath

Subsetted Swath

Geographic Masking



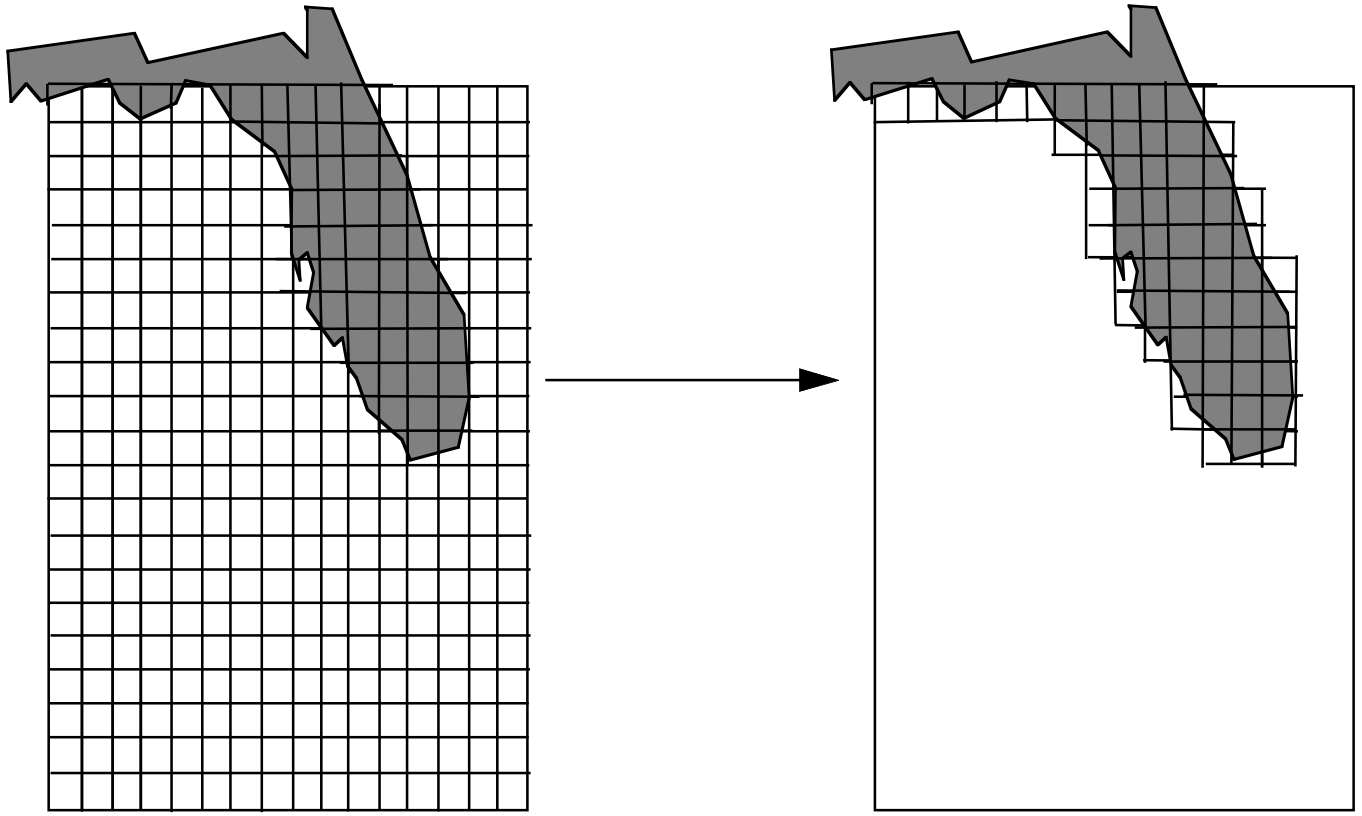
Background:

- **Certain MODIS processing at EDC will be done only for land pixels. Masking out ocean pixels and compressing the data prior to transfer will reduce the amount of data by nearly 70%. This masking technique will also be used with snow and ice masks.**

Approach:

- **ECS will provide PGEs which will apply geographical masks to data sets and insert fill values in areas which are not of interest in the subsequent processing. The masked data sets are then compressed for inter-DAAC transfer.**

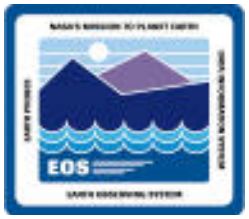
Geographic Masking



Original Data

Masked Data

Compression for Fill Values



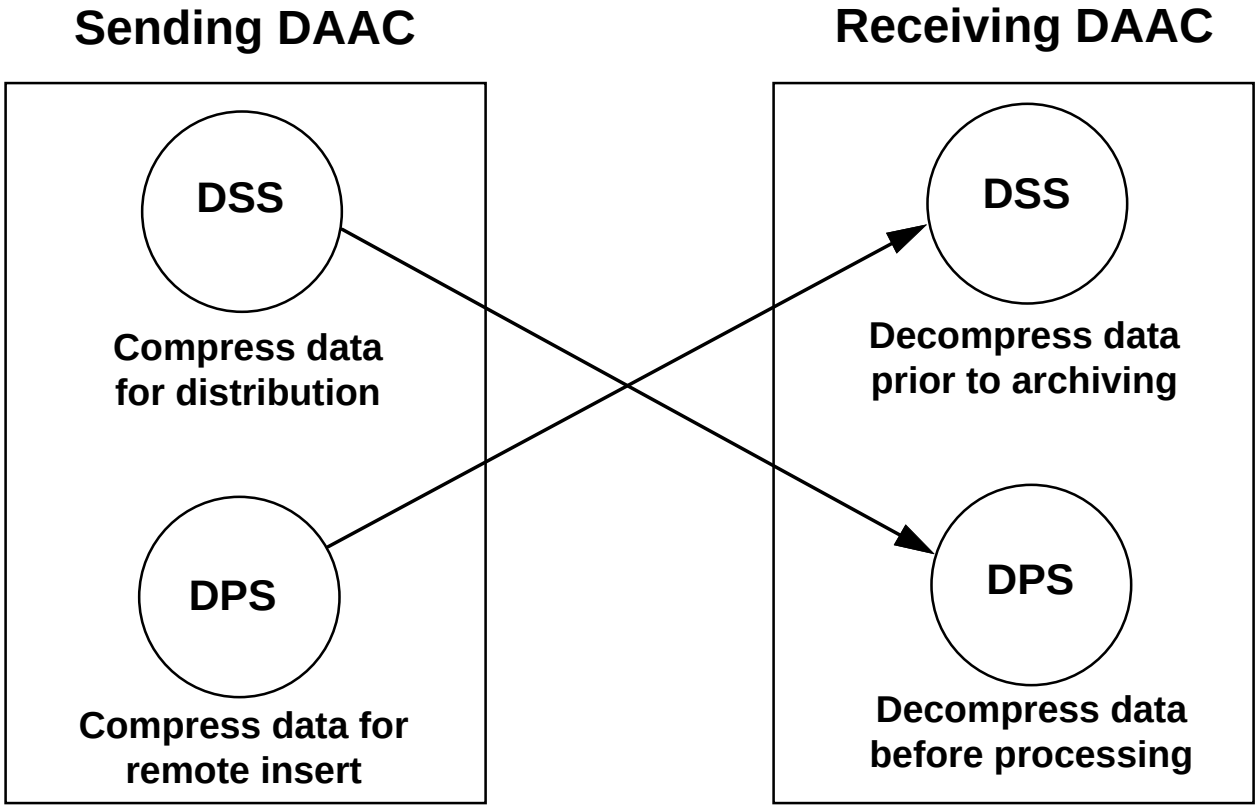
Background:

- **Some data sets already use fill values (as opposed to those requiring ECS to apply a geographic mask). Simple compression of these data sets prior to transfer will reduce network traffic.**

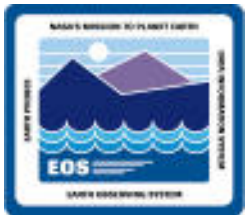
Approach:

- **Data sets will be compressed prior to inter-DAAC transfer. Compression will be performed by DPS for data to be archived at the remote site and by DSS for other data sets. Decompression will be performed by the subsystem receiving the data.**

Compression for Fill Values



Summary



- **Production Subsetting will routinely occur for a specific set of ESDTs which are produced at one DAAC and used for production and/or are archived at another DAAC.**
- **Subsetting of will reduce the daily network data transfers associated with 23 currently identified ESDTs from approximately 500 GB to 140 GB or 72%.**

Next Steps:

- **Analyze June AHWGP Baseline for changes which effect the choice of ESDTs and which techniques are to be used on them.**
- **Determine source and format of geographic masks.**